Empirical Education

Combining Multiple Measures of Teacher Effectiveness:
Preliminary Quantitative Findings

Arizona Pilot Project Summer Institute

July 22, 2013

Overview

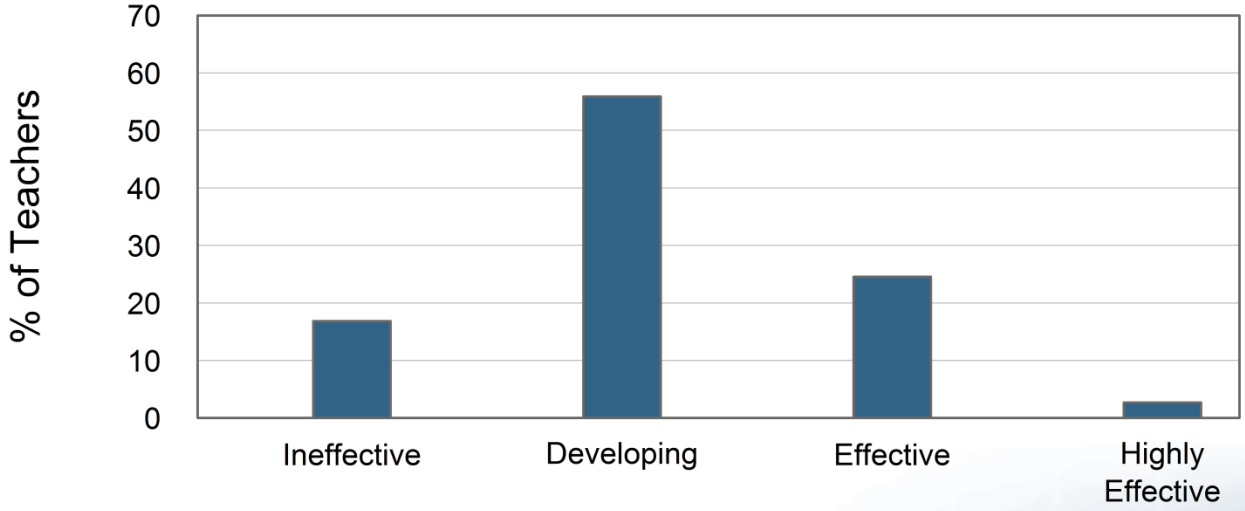
I. Analysis of Composite Performance Scores
   - Analyze scoring formulas and performance cutoff points
   - Establish whether the system treats any of the groups preferentially

II. Analysis of Component Observation, Survey and Student Academic Progress Data
   - Establish effectiveness of components in differentiating between high and low performing teachers
   - Establish correlational relationships between each component

III. Comparative Analysis of the Observational Instrument
   - Establish effectiveness and consistency of observational instrument as implemented in pilot compared to FFT data collected during the Measures of Effective Teaching project

I. Composite Teacher Performance Scores: Distribution by Classification

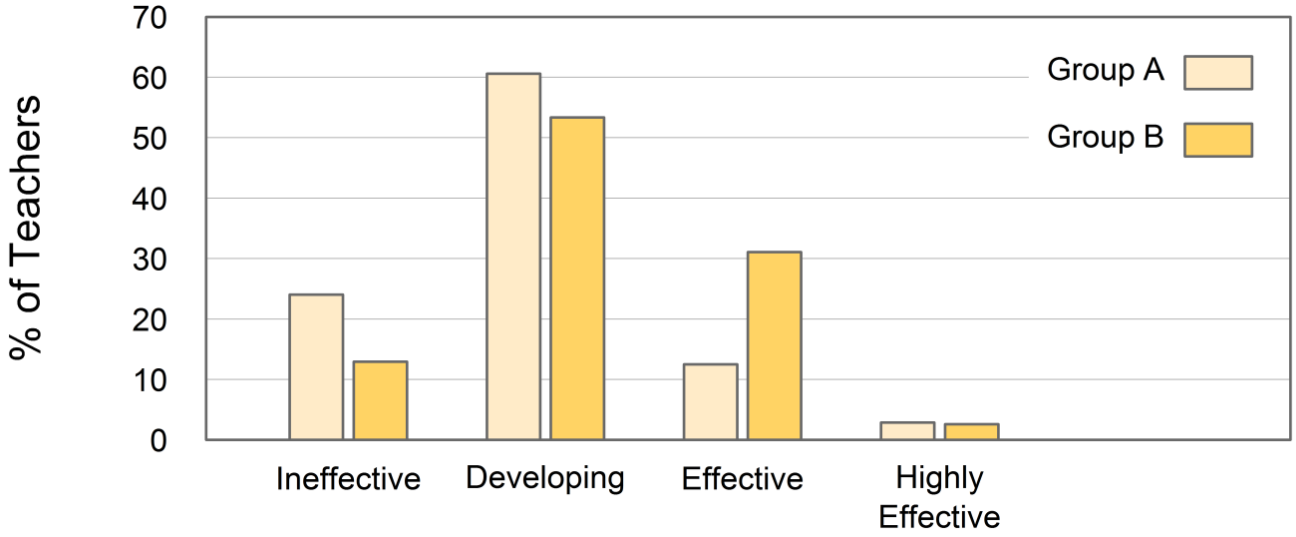
Teacher Performance Classifications	
Highly Effective	108-120 points
Effective	84-107 points
Developing	60-83 points
Ineffective	<60 points



- Large number of teachers classified as "developing" or "ineffective"
- Distribution has appearance of being "skewed" to the left

Comparison of Composite Performance Scores for Teachers With (A) and Without (B) Classroom Achievement Data

Group A = Teachers **with** available classroom-level student achievement data
Group B = Teachers **with limited or no** available classroom-level student achievement data



Group	Number of Teachers
A	104 (35%)
B	193 (65%)
Total	297 (100%)

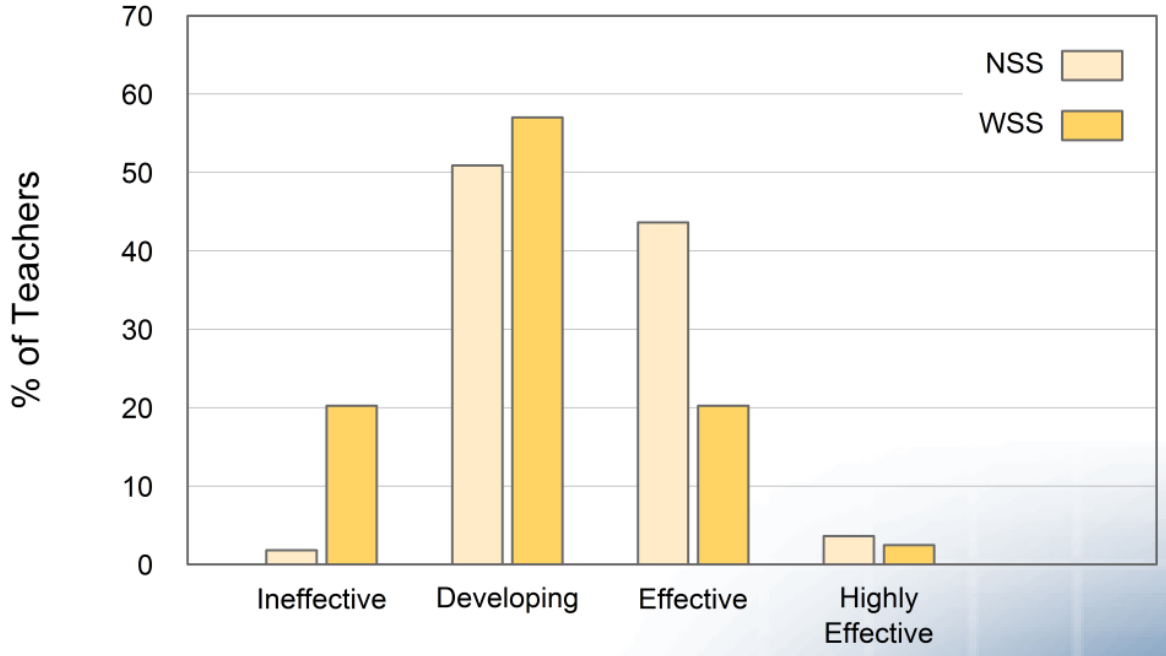
- Distribution for teacher without classroom achievement scores (B) is skewed to the right compared to teachers with such scores (A), see especially "Effective" category

Comparison of Composite Performance Scores for Teachers With-Student-Survey (WSS) vs. No-Student-Survey (NSS)

WSS = Teachers that were to have administered a student survey. Contains both A and B teachers.

NSS = Teachers that were NOT to have administered a student survey. This included teachers in K-2 grades and teachers of the SPED-Life Skills course. Contains both A and B teachers.

Group	Number of Teachers
WSS Total	242 (81%)
NSS Total	55 (19%)
Overall Total	297 (100%)

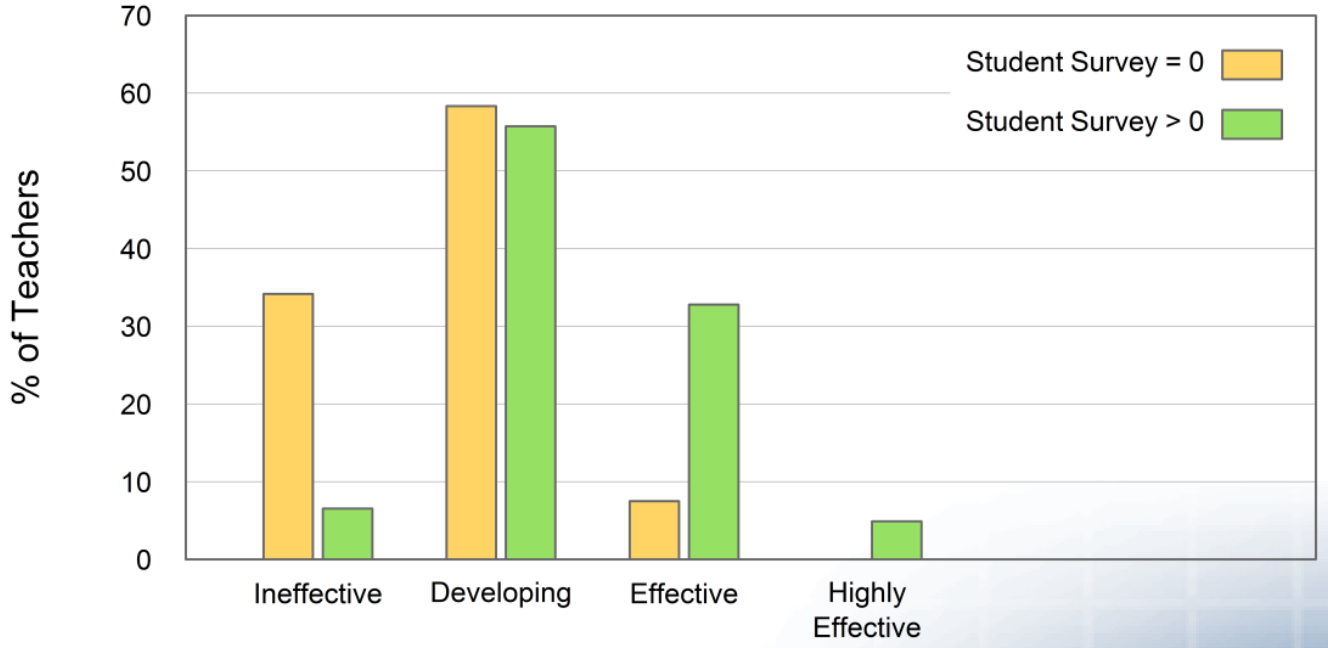


- Graph shows that WSS appears to be skewed to the left compared to NSS (see especially Ineffective and Effective categories)

Composite Performance Scores of WSS Teachers Comparing Teachers with Score of 0 to those with Score Greater than 0

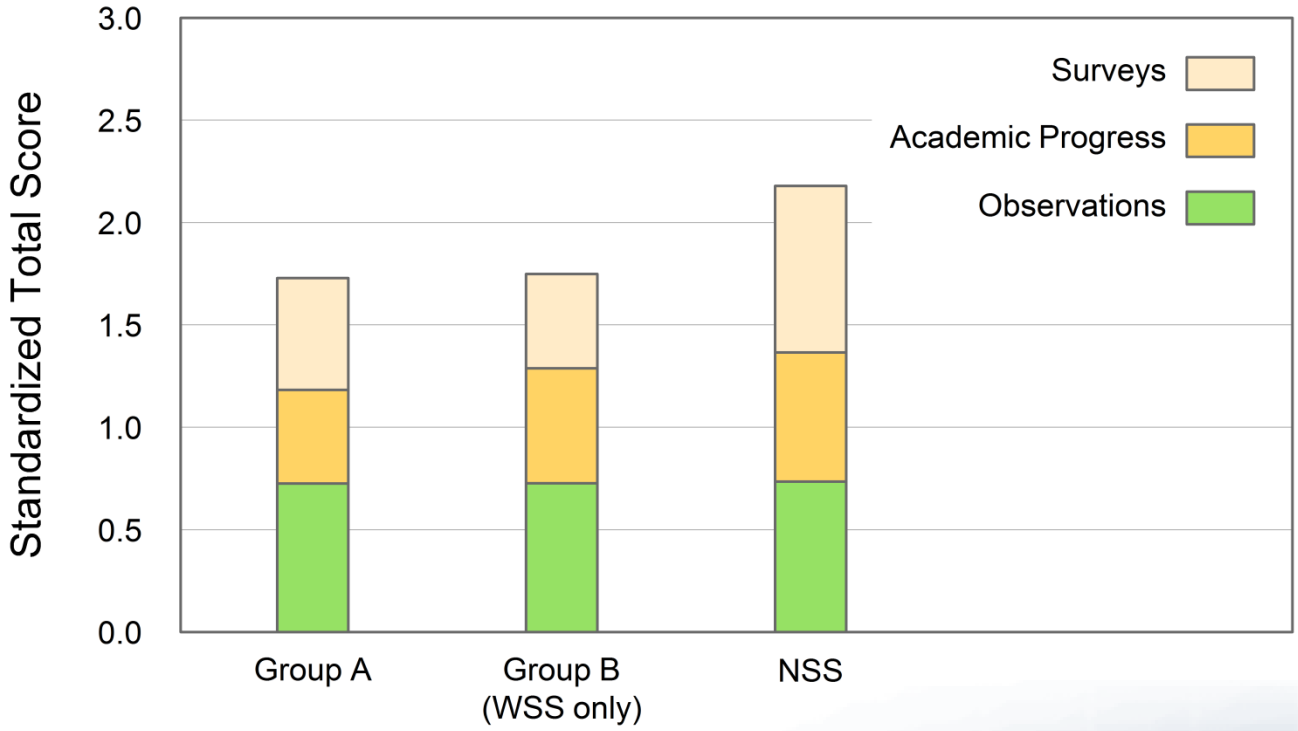
SS = 0 Teachers that were to have administered a student survey but received no score (most likely failed to administer the survey.
SS > 0 Teachers that were to have administered a student survey and received a non-zero score

Group	Number of Teachers
WSS (SS=0)	120 (40%)
WSS (SS>0)	122 (41%)
WSS Total	242 (81%)



- NSS (previous slide) is similar to WSS teachers with Student Survey > 0
- The graph shows that a student survey score of zero pushes scores down
- We found no evidence that zero student survey scores are associated with poor performance on other metrics

Summative Scores Combining the Components: Comparison of Three Groups of Teachers



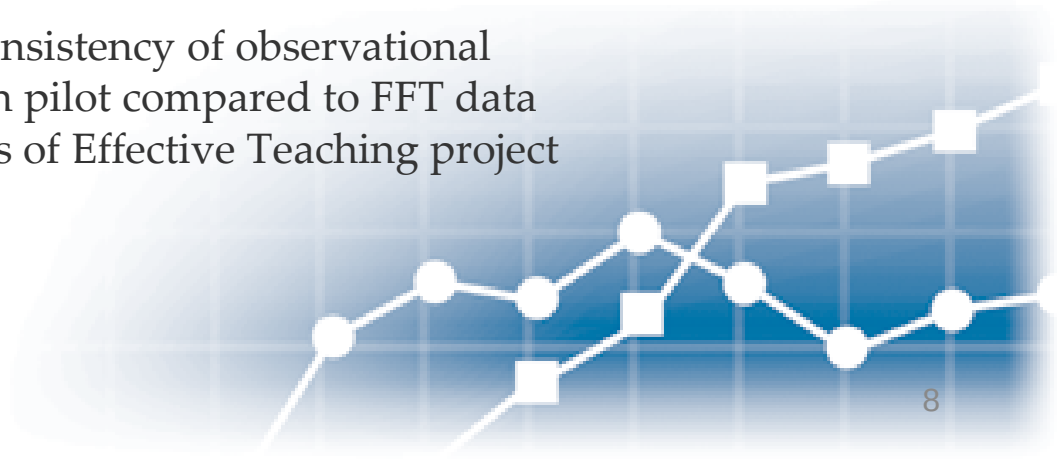
- Surveys and Academic Progress scores differ across groups
- Only observation scores are consistent across groups

Overview

I. Analysis of Composite Performance Scores
   - Analyze scoring formulas and performance cutoff points
   - Establish whether the system treats any of the groups preferentially

II. Analysis of Component Observation, Survey and Student Academic Progress Data
   - Establish effectiveness of components in differentiating between high and low performing teachers
   - Establish correlational relationships between each component

III. Comparative Analysis of the Observational Instrument
   - Establish effectiveness and consistency of observational instrument as implemented in pilot compared to FFT data collected during the Measures of Effective Teaching project



Which Components of the Composite (Student Academic Progress, Observation Domains & Surveys) Differentiate Performance Levels?

	SAP	Domain 1	Domain 2	Domain 3	Domain 4	Obs Total	Student Survey	Parent Survey	Peer Survey	Survey Total
GROUP A										
Highly Effective vs Effective	YES	YES	YES	YES	NO	YES	NO	NO	NO	YES
Effective vs Developing	YES	YES	YES	NO	NO	YES	YES	NO	NO	NO
GROUP B (WSS only)										
Highly Effective vs Effective	YES	YES	YES	YES	YES	YES	NO	YES	NO	NO
Effective vs Developing	YES	YES	YES	YES	YES	YES	YES	NO	NO	YES
NSS										
Highly Effective vs Effective	YES	YES	YES	NO	YES	YES	n/a	YES	NO	NO
Effective vs Developing	YES	YES	YES	YES	YES	YES	n/a	NO	NO	NO

- YES = Average component scores differ between the two performance levels significantly
- All components can differentiate between ineffective and developing teachers
- In general surveys tend to differentiate poorly among upper performance levels

II. Correlations Between Component Measures

Whole Group

	SAP	D1	D2	D3	D4	Obs Tot	Stud S	Par S	Peer S	Surv Tot
SAP	x									
D1		x								
D2		HI	x							
D3		HI	HI	x						
D4	LO	MOD	MOD	MOD	x					
Obs Tot	LO	HI	HI	HI	HI	x				
Stud S		LO	LO	LO	LO	LO	x			
Par S	LO		LO	LO	MOD	LO		x		
Peer S	LO				MOD	LO		MOD	x	
Surv Tot		LO	LO	LO		LO	HI*			x

- Low correlations between Acad. Prog. (SAP) and other metrics overall
- Little relationship between classroom observations and surveys
- Survey total is only a function of the student survey score (correlation < 1), but the student survey is not correlated with Academic Progress while parent and peer surveys are
- Student survey not correlated with Academic Progress but this could be due to a large number of zero values

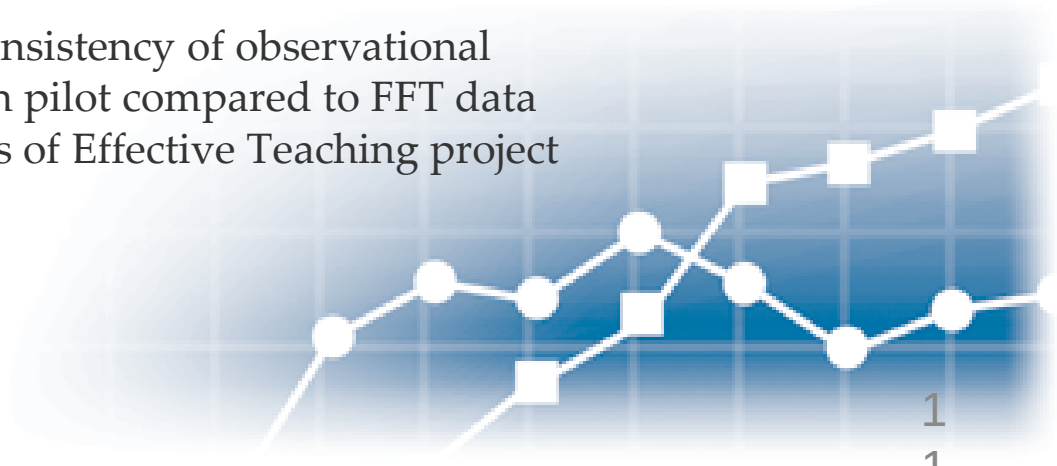
*	Nearly perfect correlation
High	.70 - .99
Moderate	.45 - .66
Low	.14 - .33
Blank cell	Correlation not significantly different from zero

Overview

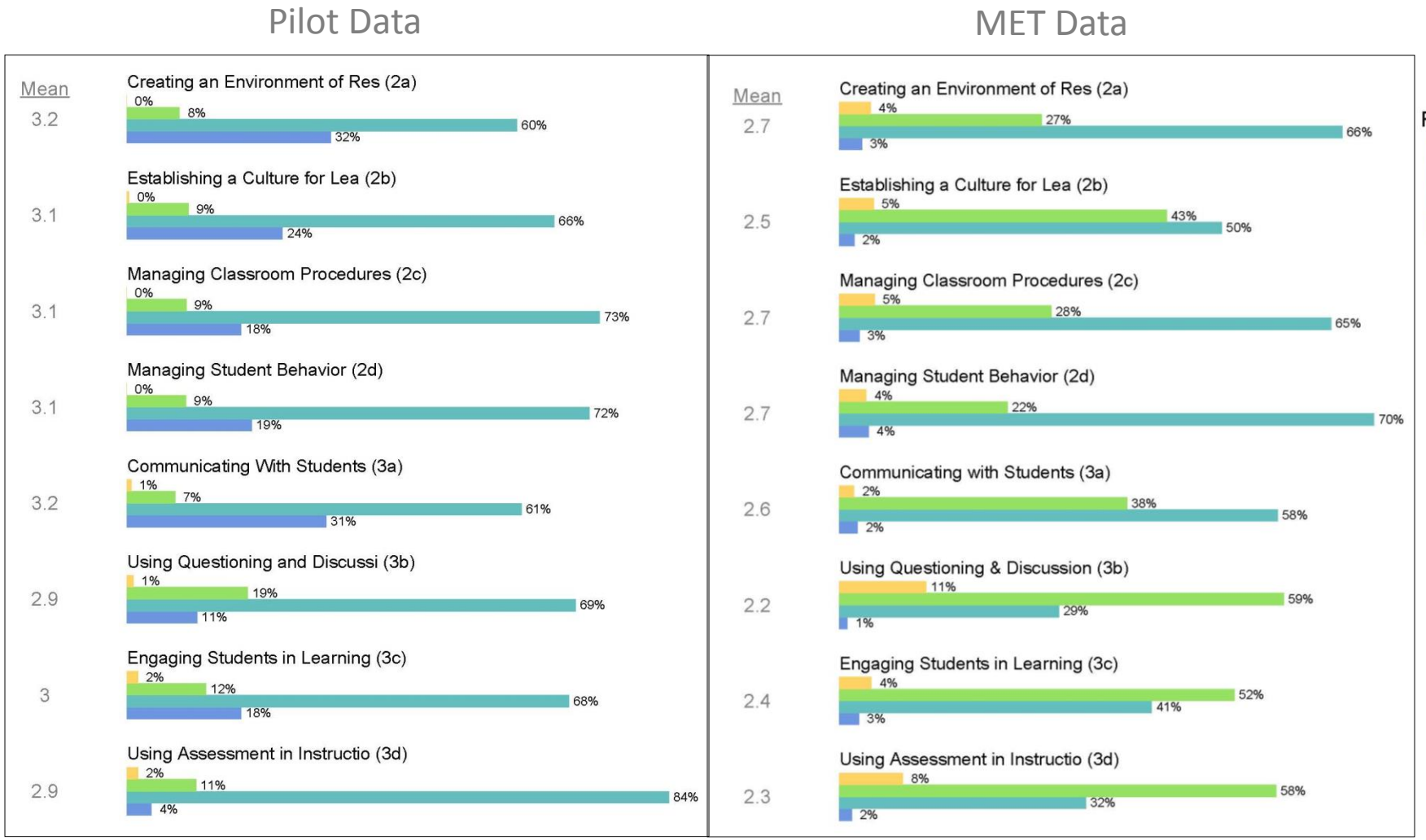
I. Analysis of Composite Performance Scores
   - Analyze scoring formulas and performance cutoff points
   - Establish whether the system treats any of the groups preferentially

II. Analysis of Component Observation, Survey and Student Academic Progress Data
   - Establish effectiveness of components in differentiating between high and low performing teachers
   - Establish correlational relationships between each component

III. Comparative Analysis of the Observational Instrument
   - Establish effectiveness and consistency of observational instrument as implemented in pilot compared to FFT data collected during the Measures of Effective Teaching project



III. Distribution of Ratings by Item of the Observational Instrument



- Compared to the MET project, observation scores are less dispersed and biased upward.
- Possible problem with implementation of FFT/rater training or with the process for producing final observation scores from two rounds of live observation

Empirical Education

Preliminary Results Summary

I. Analysis of Composite Performance Scores

– Most teachers are rated less than "Effective"

– A 0 value for student surveys skewed scores lower

– Some inconsistency in how different groups are evaluated

II. Analysis of Component Observation, Survey and Student Academic Progress Data

- Survey scores are inconsistent – large number of zero values for Student Survey
- Little consistency between classroom observations and other components

III. Item-Level Analysis of the Observational Instrument

– Appears to be biased upward and with lower variability compared to MET